

Remarks on Category-Based Routing in Social Networks

Karl Bringmann, Kurt Mehlhorn, and Adrian Neumann

Max-Planck-Institut für Informatik

Abstract It is well known that individuals can route messages on short paths through social networks, given only simple information about the target and using only local knowledge about the topology. Sociologists conjecture that the routes are found greedily by passing the message to an acquaintance that has more in common with the target than the current holder of the message, e.g. if a dentist in Saarbrücken wants to send a message to a specific lawyer in Munich, he may forward it to someone who is a lawyer and/or lives in Munich. Modelling this setting, Eppstein et al. introduced the notion of *category-based routing*. The goal is to assign a set of categories to each node of a graph such that the categories are connected, greedy routing is possible, and the maximal number of categories a node is in is minimal. We give an improved upper bound construction and show lower bounds for stars, trees, and general graphs.

1 Introduction

In the 1960's, Milgram [1–3] observed the *small world phenomenon*, i.e. that short paths seem to connect us all in the social graph in which the vertices represent persons and two persons are connected by an edge if and only if they know each other. He performed experiments in which he asked randomly selected participants to relay letters across the USA by passing them to one of their direct acquaintances.

The participants only had simple categorical information about the target, such as name, location, and occupation, and knowledge about their own acquaintances, that is, they knew the local topology of the network. The messages arrived typically after only six hops.

Perhaps even more surprising than the mere existence of short paths in social networks is the apparent ease with which humans can discover them, despite having only little information. Experiments by sociologists [4–7] indicate that we use a simple greedy heuristic to route. The message is passed on to the acquaintance that is most similar to the target.

In [8], Eppstein et al. model this setting as a graph $G = (V, E)$, where people act as vertices and whose edges represent pairwise acquaintance, together with a system of categories $\mathcal{S} \subseteq \mathcal{P}(V)$. Each category $C \in \mathcal{S}$ is the vertex set of a connected subgraph of G . For two vertices s and t , let the distance from s to t be the number of categories containing t but not s , i.e.,

$$d(s, t) = |\{C \in \mathcal{S} \mid t \in C \text{ and } s \notin C\}|.$$

A system of categories \mathcal{S} supports *greedy routing* in G (is *good* for G) if for any two vertices s and t , there is a neighbor u of s with $d(u, t) < d(s, t)$. For a system of categories \mathcal{S} , Eppstein et al. define its *membership dimension* as the maximal number of categories to which any vertex belongs:

$$\text{memd}(\mathcal{S}) := \max_{v \in V} |\{C \in \mathcal{S} \mid v \in C\}|.$$

The goal is then to construct a system of categories supporting greedy routing and having small membership dimension. Membership dimension captures the cognitive load of the participants, i. e. the number of categories an actor must keep track of in order to decide on the next node of the route.

Eppstein et al. [8] show the existence of a good system of categories \mathcal{S} with

$$\text{memd}(\mathcal{S}) \in \mathcal{O}((\text{diam}(G) + \log |V|)^2)$$

and note a lower bound of $\text{diam}(G)$; here $\text{diam}(G)$ denotes the diameter of G .

We substantially improve on the upper bound stated above and establish new lower bounds. We review related work in Sect. 2, introduce notation in Sect. 3, and prove exact bounds for lines, grids, and tori in Sect. 4. In Sect. 5, we construct for every graph G a good system of categories \mathcal{S} with

$$\text{memd}(\mathcal{S}) = \mathcal{O}\left(d \cdot \log\left(\frac{2|V|}{d}\right)\right),$$

where $d = \text{diam}(G)$. This bound improves upon the bound of Eppstein et al. except for $\text{diam}(G) = \Theta(\log |V|)$. In Sect. 6 we show an almost matching lower bound. We exhibit for all nonnegative integers n and d a graph with $1+nd$ vertices and diameter $2d$ for which every

good system of categories has membership dimension

$$\Omega\left(\frac{d \ln(|V|/d)}{\ln(d \ln(|V|/d))}\right).$$

In Sect. 7 we show that every good system \mathcal{S} for a graph G with average degree δ has membership dimension

$$\Omega\left(\text{diam}(G) + \frac{\log |V|}{\log \delta}\right),$$

in particular, bounded degree graphs require logarithmic membership dimension. The bound is best possible. For each triple (n, δ, diam) of positive reals with $1 \leq \delta, \text{diam} \leq n$ we exhibit a graph with $\Theta(n)$ vertices, average degree $\Theta(\delta)$, and diameter $\Theta(\text{diam})$ for which a good system of membership dimension

$$\mathcal{O}\left(\text{diam} + \frac{\log |V|}{\log \delta}\right)$$

exists.

2 Related Work

We essentially paraphrase the analogous section in [8].

Greedy Routing is a well-studied technique with many applications. A variety of methods are known, for example geographical information as an aid for routing has been explored [9, 10]. This method does not succeed on all networks, so a number of enhancements have been developed [11–13]. Several groups also examine *succinct* greedy-routing strategies that limit the additional information at every vertex to be logarithmic in the size of the network [14–17]. Category-based routing with logarithmic size membership dimension is a special case of succinct greedy routing.

A category-based approach has been studied by Mei et al. [18] as a heuristic for routing in dynamic networks. Nodes are assumed to be organized into categories by the user. Experimental evidence indicates that category-based routing outperforms heuristics based on location or random choices. Eppstein et al. [8] were the first who studied it from a complexity-theoretic viewpoint. We reviewed their results in the introduction.

A different approach to routing in social networks was studied by Kleinberg [19]. He focuses on location instead of categorical information to explain how we find short routes efficiently. Based on his insights he constructs a random graph model that has similar properties and shows for which parameters routing is possible. In contrast the approach of Eppstein et al. seeks to construct a system of categories that enables greedy routing for a given network.

3 Preliminaries

Unless stated otherwise, we consider undirected graphs $G = (V, E)$ with $n = |V|$ nodes. For two nodes $u \in V, v \in V$ let $\text{dist}(u, v)$ be the number of edges on a shortest path between u and v . Then define the diameter of a graph G as

$$\text{diam}(G) := \max_{u \in V, v \in V} \text{dist}(u, v),$$

that is, the diameter is the length of a longest shortest path in G .

Let $\mathcal{S} \subseteq \mathcal{P}(V)$ be a system of subsets of the vertices of G that induce connected subgraphs. For a node $u \in U$ define $\text{cat}(u)$ to be the set of categories to which u belongs

$$\text{cat}(u) := \{C \in \mathcal{S} \mid u \in C\}.$$

The membership dimension $\text{memd}(\mathcal{S})$ is the maximum number of categories to which any node belongs, that is,

$$\text{memd}(\mathcal{S}) := \max_{u \in V} |\text{cat}(u)|.$$

We say a system of categories supports greedy routing in G or is *good* for G if a greedy routing algorithm succeeds. If the message addressed to node $t \in V$ is currently in node $u \in V$ and $u \neq t$, the algorithm forwards it to a neighbor $v \in N(u)$ of u that is closer to t according to the distance function

$$d(v, t) := |\text{cat}(t) \setminus \text{cat}(v)|,$$

The algorithm succeeds if for all $u \in V$ there is a neighbor $v \in N(u)$ such that $d(v, t) < d(u, t)$.

4 Simple Bounds

Already in Eppstein et al. [8], it is observed that the diameter bounds the membership dimension from below.

Lemma 1. *For any graph G and good system of categories \mathcal{S} we have*

$$\text{memd}(\mathcal{S}) \geq \text{diam}(G).$$

Proof. See Appendix, Proof 1.

For paths this bound is tight, cf. the construction in Fig. 1. We can extend this observation to all graphs that can be obtained from paths and cycles by taking cross products.

Definition 1. *Let $G = (V, E)$, and $H = (V', E')$, be two graphs. Then the cross product $G \times H$ is the graph (\tilde{V}, \tilde{E}) with*

$$\begin{aligned} \tilde{V} &= V \times V', \\ \tilde{E} &= \{(u, x), (v, y)\} \mid \\ &\quad (\{u, v\} \in E \wedge x = y) \vee \\ &\quad (\{x, y\} \in E' \wedge u = v)\}. \end{aligned}$$



Figure 1. Minimal categories for routing from the leftmost vertex to the rightmost. Adding the symmetric categories allows to route between any pair of vertices. Every vertex is contained in exactly four categories.

Lemma 2. *Let $M(G)$ be the minimal membership dimension needed to route in G . Then*

$$M(G \times H) \leq M(G) + M(H).$$

Proof. See Appendix, Proof 2.

Lemmas 1 and 2 immediately give the following tight bounds.

Corollary 1. *For grid graphs G we get $M(G) = \text{diam}(G)$. \square*

Corollary 2. *For hypercubes G we get $M(G) = \text{diam}(G) = \lceil \log n \rceil$. \square*

Corollary 3. *For tori $G = C_k \times C_l$ we have $M(G) = \lceil k/2 \rceil + \lceil l/2 \rceil$. \square*

5 Improved Upper Bound

We construct for every graph G a good system of categories \mathcal{S} with membership dimension

$$\mathcal{O} \left(\text{diam}(G) \log \frac{|V|}{\text{diam}(G)} \right).$$

It suffices to prove the bound for trees. For general graphs, we construct a spanning tree of diameter $\text{diam}(G)$ and route in the spanning tree.

Lemma 3. *For any tree on n nodes with diameter d there is a system of categories \mathcal{S} of membership dimension*

$$\text{memd}(\mathcal{S}) = \mathcal{O}(d \log(2n/d)).$$

We conjecture this bound to be tight. An example might be the star with degree n/d sending out paths of length d as shown in Fig. 6.

Proof. For a tree T let $r \in V(T)$ and consider a triple (r, L, R) with L, R a partitioning of the neighbors of r in T . After deleting r , the tree falls into components; take the ones containing nodes from L . To these components add r again (making it a neighbor to all nodes in L) to get a tree $T_L = T_L(r)$. Build T_R symmetrically. This cuts T into two trees; each vertex is in exactly one of T_L and T_R except for r , which belongs to both trees.

Now, we call (r, L, R) a *balanced routing cut* if both $|V(T_L)|$ and $|V(T_R)|$ are at least $c_1 \cdot |V(T)|$ and at most $c_2 \cdot |V(T)|$. A basic graph theoretic argument shows that any tree has a balanced routing cut for constants $c_1 = 1/3$ and $c_2 = 2/3$, see for example [20].

Given a tree T we now construct categories as follows. Take a balanced routing cut (r, L, R) of T . Similar to Eppstein et al. [8], we construct categories that allow routing from any vertex in T_R all the way to r when having a target in T_L (and symmetrically), see Fig. 2. Then we recurse on T_R and T_L . After that, we unify some categories to decrease the membership dimension.¹

The base case of this procedure is a graph of constant size, where we add any valid system of categories of constant size. Observe that this way we construct a valid system of categories for T .

¹ The recursion means that we again find a balanced routing cut in T_R , so the vertex we split at in T_R does not have to be r or in R . This is a major difference to Eppstein et al.

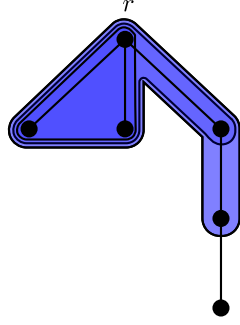


Figure 2. Categories that allow routing from T_R to r for a target in T_L .

We now describe this construction in more detail. Let $d_v = \text{dist}(v, r)$ be the distance from $v \in V(T)$ to r in T .² For routing from T_R to T_L we add the categories $V(T_L) \cup \{v \in V(T_R) \mid d_v \leq k\}$, for $0 \leq k \leq \text{diam}(T_R)$.³ We add symmetric categories for routing from T_L to T_R . Note that these categories allow us to route from any vertex in T_R to r when having a target in T_L .

Then we recurse on T_R and T_L .

After that, we change the system of categories slightly to decrease its membership dimension. We have split at (r, L, R) and in the recursion on T_R we constructed, say, categories R_1, \dots, R_a containing r and in the recursive call to T_L we constructed categories L_1, \dots, L_b containing r . Assume wlog. $a \leq b$. Then we can replace the R_i and L_j by the categories

$$\{R_i \cup L_i \mid 1 \leq i \leq a\} \cup \{L_{a+1}, \dots, L_b\}.$$

These categories are still connected and greedy routing is still possible, as the R_i (L_j) were only needed to route inside T_R (T_L). See Fig. 3.

² or in T_R or T_L , there is no difference if u is in that tree.

³ This is the same construction as in Eppstein et al. [8]

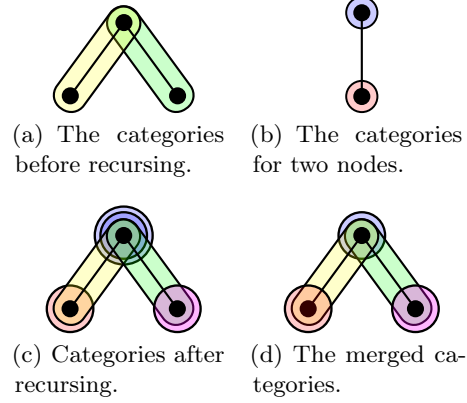


Figure 3. Merging of categories after recursing.

This ends the construction of categories. Observe that we construct a valid system of categories.

It remains to bound the membership dimension of the constructed system of categories. Consider any node $v \in V(T)$. Joining categories as described in the third part of our construction implies that the number of categories containing v equals the maximal number of categories containing v constructed on any path in the recursion tree. As we select balanced routing cuts, in each recursive call of the construction the number of nodes of the subproblems is decreased by a constant factor and hence each such path has length $\mathcal{O}(\log(n))$.

Consider one such path. On level i , $0 \leq i \leq \mathcal{O}(\log(n))$, we considered a subproblem T_i on at most $n \cdot c_2^i$ nodes. The diameter of T_i is bounded from above by $d = \text{diam}(T)$, as T_i is a subtree of T . Moreover, the diameter of T_i is bounded by its number of nodes. Hence,

$$\text{diam}(T_i) \leq \min\{d, n \cdot c_2^i\}.$$

Our procedure cuts T_i and adds some categories to allow routing over that

cut. Observe that the number of such categories is bounded by $\mathcal{O}(\text{diam}(T_i))$. Thus, the number of categories containing v can be bounded from above by (omitting the $\mathcal{O}()$)

$$\sum_{i=0}^{\mathcal{O}(\log(n))} \min\{d, n \cdot c_2^i\}.$$

We can bound this sum as

$$\sum_{i=0}^{-\log_{c_2}(n/d)} d + \sum_{i=-\log_{c_2}(n/d)+1}^{\mathcal{O}(\log(n))} n \cdot c_2^i,$$

which simplifies to

$$d \log(n/d) + d = \mathcal{O}(d \log(2n/d)).$$

□

6 Stars

In this section we prove upper and lower bounds for stars. The lower bound for stars nearly matches the upper bound of the preceding section. A star of diameter $2d$ and ℓ leaves has $1+\ell d$ nodes. The center node has degree ℓ and each leaf is joined to the center node by a path of length d . Fig. 4 shows a star with four leaves and diameter 2. We start with simple upper and lower bounds.

Lemma 4. *In a star with ℓ leaves, the center node is contained in at least $\log \ell$ categories.*

Proof. For every ordered pair (i, j) with $i \neq j$, there must be a category C containing j and the center, but not i . Otherwise, every category containing the center and j also contains i and hence the set of categories containing $\{j, c\}$ is the same as the set of categories containing $\{i, j, c\}$. Thus the distance from

i to j does not decrease along the edge ic , a contradiction. We conclude that the number of categories containing the center is at least $\log \ell$.

An alternative proof can be found in the Appendix, Proof 3. □

The same argument establishes:

Lemma 5. *In a tree, each node v is contained in least $\log \deg v$ categories.* □

Lemma 6. *For a star with ℓ leaves and diameter 2, there is a system of categories \mathcal{S} with membership dimension*

$$\text{memd}(\mathcal{S}) \leq 1 + 2 \lceil \log \ell \rceil.$$

Proof. Let the leaves be numbered from 1 to ℓ . Every leaf forms a category of its own. For every i , $0 \leq i < \lceil \log \ell \rceil$ we have two categories Z_i and O_i : Z_i contains the center and all leaves that have a zero in the i -th bit of their binary representation; O_i contains the center and all leaves that have a one in the i -th bit of their binary representation. See Fig. 4. Clearly every node is contained in at most $1 + 2 \lceil \log \ell \rceil$ categories.

Consider any two leaves u and v and assume we want to route from u to v . Let k be the number of positions in which the binary representation of u and v differ. Then $d(u, v) = 1 + k \geq 2$ and $d(c, v) = 1$. Thus we can successfully route from u to v . □

We next improve on both bounds.

Lemma 7. *Let k be minimal such that*

$$\binom{k}{\lfloor k/2 \rfloor} \geq \ell.$$

Then $k = \log n + (\frac{1}{2} + o(1)) \log \log n$.

There is a system of categories \mathcal{S} that supports routing in stars with n leaves and diameter 2 having

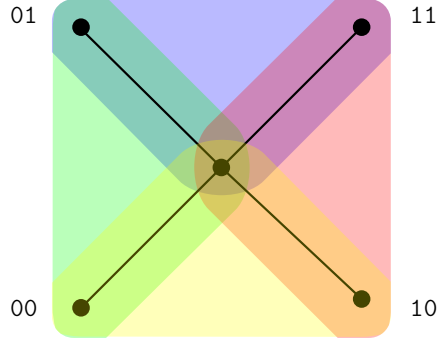


Figure 4. Categories Z_i and O_i for a star with $\ell = 4$.

$$\text{memd}(\mathcal{S}) = k.$$

This bound is tight, i.e., every good system of categories \mathcal{S} for this graph has

$$\text{memd}(\mathcal{S}) \geq k.$$

Proof. See Appendix, Proof 4.

Similar techniques apply for stars with larger diameters.

Lemma 8. *For any star with ℓ leaves, in which every leaf is connected by a path of length d to the center, there is a system of categories \mathcal{S} with membership dimension*

$$\text{memd}(\mathcal{S}) \in \mathcal{O}(d \log \ell).$$

Proof. The Lemma follows from an application of the algorithm from Lemma 3. For an alternative proof see Appendix, Proof 5. \square

We proceed to show a lower bound. A key ingredient of our proof is the following problem of covering the set

$$Q = \{(i, j) \mid 1 \leq i, j \leq \ell, i \neq j\}$$

by t rectangles, i.e.,

$$Q = \bigcup_{1 \leq k \leq t} S_k \times T_k.$$

Clearly $S_k \cap T_k = \emptyset$ in any covering of Q by rectangles. We will first prove a lower bound for the rectangle covering problem and then obtain a lower bound for membership dimension by a reduction to the rectangle covering problem. The intuition for the reduction is that every category C_k corresponds to a set $S_k \times T_k$ such that $(i, j) \in S_k \times T_k$ if C_k allows us to route between i and the center when having j as target. Hence by bounding the number of rectangles needed to cover Q , we bound the number of categories that need to contain the center in a star with n leaves.

Lemma 9. *If $4 \ln t \leq \ln \ell$, we have*

$$\sum_k |S_k| \geq \frac{\ell \ln \ell}{32 \ln t}$$

in any covering of Q by t rectangles.

Proof. See Appendix, Proof 6.

We now consider a star with ℓ leaves in which the arms have length d . Let

$$\mathcal{P} = \{P^{(1)}, \dots, P^{(t)}\}$$

be the family of t categories containing the center. We will lower bound t using the results from Lemma 9.

Every category containing the center corresponds to a partition

$$P = (P_0, P_1, \dots, P_d)$$

of $[0, \dots, \ell]$, where P_k contains the leaves for which a path of length k from the center to the leaf is contained in the category.

Lemma 10. *For every i and j with $1 \leq i \neq j \leq \ell$ and every k , $0 \leq k < d$, there must be a $P \in \mathcal{P}$ such that*

$$(i \in P_k) \neq (j \in P_k).$$

Proof. There must be a P such that $i \in P_k$ and $j \in P_d$ as otherwise we cannot route from i to j . Since P is a partition, $j \notin P_k$. \square

Lemma 11. *Let t be the number of categories containing the center in a star with ℓ leaves in which every leaf is connected to the center by a path of length d . Then*

$$t \ln t \geq \frac{d \ln \ell}{32}.$$

Proof. This proof uses a double counting technique. For each $1 \leq i \leq \ell$ and k , define the vector $v_{i,k}$ by

$$(v_{i,k})_j = (i \in (P^{(j)})_k).$$

Then by Lemma 10 for every k , the vectors in $\{v_{i,k} \mid 1 \leq i \leq \ell\}$ are pairwise distinct and hence by Lemma 9 for every k , $0 \leq k < d$,

$$\sum_{1 \leq j \leq t} |(P^{(j)})_k| \geq \frac{\ell \ln \ell}{32 \ln t}.$$

Summation over k yields

$$\begin{aligned} d \frac{\ell \ln \ell}{32 \ln t} &\leq \sum_{k=0}^{d-1} \sum_{j=1}^t |(P^{(j)})_k| \\ &= \sum_{j=1}^t \sum_{k=0}^{d-1} |(P^{(j)})_k| \\ &\leq t\ell. \end{aligned} \quad \square$$

Theorem 1. *Consider a star in which each of the ℓ leaves is connected to the center by a path of length d and let t be the number of categories that contain the center. If $\ell \geq 3$,*

$$t \geq \frac{d \ln \ell}{32(\ln d + \ln \ln \ell)}.$$

Proof. Assume otherwise, and let $X = d \ln \ell$. Since $t \ln t$ is an increasing function in t , Lemma 11 implies

$$\frac{X}{32 \ln X} \ln \left(\frac{X}{32 \ln X} \right) \geq \frac{X}{32}$$

and hence

$$\ln \left(\frac{X}{32 \ln X} \right) \geq \ln X,$$

a contradiction. \square

7 A Universal Lower Bound

The lower bounds of the preceding section are existential. We showed the existence of graphs for which every good system of categories has a certain membership dimension. The lower bound almost matches the universal upper bound of Sect. 5. In this section, we show a universal lower bound: every good system of categories for a graph G has membership dimension

$$\Omega \left(\text{diam}(G) + \frac{\log |V|}{\log \delta} \right),$$

where δ is the average degree. We also show that this bound is best possible.

Theorem 2. *In a graph $G = (V, E)$ with average degree $\delta/2$, there is a node that is contained in*

$$\Omega(\text{diam}(D) + \log n / \log \delta)$$

categories.

Proof. The lower bound of $\text{diam}(D)$ was already established by Eppstein et al. We turn to the second bound. Since G has average degree $\delta/2$, there are at least $n/2$ nodes that have degree less than δ . Among these we can greedily

find an independent set I of size at least $\Omega(n/\delta)$. Consider the graph

$$G' = \left(V(G), E(G) \cup \binom{V(G) \setminus I}{2} \right),$$

i. e., G where the subgraph outside of I is augmented to a clique. Routing in G' can only be easier than in G . We show a lower bound for G' .

We want to show that either a node in I or a node in its neighborhood

$$N(I) = \{u \mid v \in I \wedge \{u, v\} \in E\}$$

is in $\Omega(\log n / \log \delta)$ categories. Fix a system of categories $(C_i)_{1 \leq i \leq m}$ that allows greedy routing. For every node v in $I \cup N(I)$ define a pattern $p^{(v)}$ as

$$\left(p^{(v)} \right)_i = \begin{cases} 1 & v \in C_i \\ 0 & v \notin C_i \wedge N(v) \cap C_i \neq \emptyset \\ * & v \notin C_i \wedge N(v) \cap C_i = \emptyset. \end{cases}$$

We say two patterns $p^{(v)}, p^{(u)}$ *match* if they agree on all positions where both have a 1 or 0 (i. e., * matches to anything). Routing is only possible if no two patterns match. Let u and v be distinct vertices. Greedy routing from u to v requires the existence of a neighbor z of u and a category C_i with $C_i \in (\text{cat}(v) \cap \text{cat}(z)) \setminus \text{cat}(u)$. Then $p_i^{(v)} = 1$ and $p_i^{(u)} = 0$.

To each pattern we assign a region of points in the hypercube $\{0, 1\}^m$, namely the set of all matching bitstrings. Observe that two patterns match iff their regions in the hypercube overlap. Intuitively, this means that for allowing greedy routing we need to set many values in the $p^{(v)}$ to 0 or 1 to make these regions small enough to accommodate all without overlap. A high number of 1's forces nodes from I into many categories, a high number of 0's forces nodes from $N(I)$ into many categories.

To make this into a formal argument, consider the cost of a vertex $v \in I$ defined as

$$c(v) := t_1^{(v)} + \frac{1}{\delta} t_0^{(v)},$$

where $t_k^{(v)}$ is the number of positions in the vector $p^{(v)}$ equal to k . Note that $c(v)/2$ is a lower bound for

$$\max_{u \in \{v\} \cup N(v)} |\text{cat}(u)|,$$

the maximal number of categories u or one of its neighbors is in, since each 1 means an additional category for v and each 0 an additional category for one of its at most δ neighbors. In the remainder of the proof we show that there is a vertex $v \in I$ with $c(v) = \Omega(\log n / \log \delta)$.

For $\delta = 1$, an easy argument shows an $\Omega(\log n)$ lower bound. Define a measure on the hypercube $\{0, 1\}^m$ as

$$\mu(x) = 2^{-m},$$

for $x \in \{0, 1\}^m$, and by

$$\mu(X) = \sum_{x \in X} \mu(x)$$

for $X \subset \{0, 1\}^m$. Then the total measure of the hypercube is 1. As the regions defined by the patterns $p^{(v)}$ with $v \in I$ must be disjoint, there must be a pattern that has measure $\mu(p^{(v)})$ no more than $1/n$. As $\mu(p^{(v)}) = 2^{-c(v)}$ we get $c(v) \geq \log n$.

The same argument can be applied if δ is bounded by some constant, as then replacing δ by 1 in the definition of $c(v)$ does not change it asymptotically.

For δ greater than some large enough constant we change the measure on the hypercube to

$$\hat{\mu}(x) = \alpha^{\sum x} \cdot (1 - \alpha)^{m - \sum x},$$

for an $0 < \alpha < 1$ that is to be determined, where $\sum x$ denotes the number of 1's in the bitstring $x \in \{0, 1\}^m$ (and, thus, $m - \sum x$ the number of 0's). Again we sum up for subsets of the hypercube and the whole hypercube has measure 1. The parameter α will incorporate the reduced weight of 0's in the patterns. We want to choose α such that for some $C > 0$

$$\begin{aligned}\alpha &= 2^{-C} \\ (1 - \alpha) &= 2^{-C/\delta},\end{aligned}$$

or, equivalently, we want to find a C such that

$$1 = 2^{-C} + 2^{-C/\delta}, \quad (1)$$

as then again $\mu(p^{(v)}) = 2^{-C \cdot c(v)}$. Now, we argue as in the $\delta = 1$ case. Since the patterns may not overlap, there has to be a pattern with measure $\mu(p^{(v)})$ at most $1/n$. As $\mu(p^{(v)}) = 2^{-C \cdot c(v)}$ we get $c(v) \geq \log n / C$, which shows the claim assuming $C = O(\log \delta)$. It remains to show the latter.

Unfortunately, we cannot solve (1) exactly for C . However, for $C = O(\delta)$ we have $2^{-C/\delta} = 1 - \Theta(C/\delta)$, and, thus, for $C = \log \delta$

$$2^{-C} + 2^{-C/\delta} = \frac{1 - \Theta(\log \delta)}{\delta} < 1,$$

for δ large enough. On the other hand, for $C = \log \delta - 2 \log \log \delta$ we have

$$2^{-C} + 2^{-C/\delta} = \frac{\log^2 \delta - \Theta(\log \delta)}{\delta} > 1,$$

for δ large enough. Hence, there is a root $C = \log \delta - \Theta(\log \log \delta)$. \square

This bound is tight. For any triple (n, δ, diam) of positive reals with $1 \leq \delta, \text{diam} \leq n$, we can construct a tight

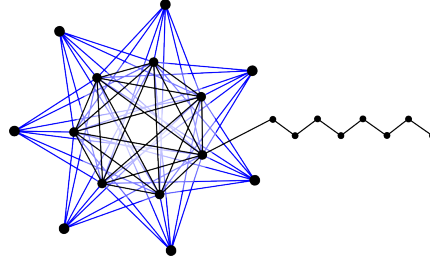


Figure 5. Tight example for Theorem 2

example with $\Theta(n)$ vertices, average degree $\Theta(\delta)$, and diameter $\Theta(\text{diam})$. For this construction take a δ -Clique K_δ and connect a set O of n additional nodes to all nodes in K_δ . Add a path P of length diam . For an example see Fig. 5. The graph G thus constructed has the required parameters.

We construct a good system of categories for G with a membership dimension matching the lower bound of $\Theta(\text{diam}(G) + \log(n/\delta))$ from Theorem 2. Routing from the nodes on P to all other nodes is possible with $\text{diam}(G)$ membership dimension as in the solution for simple paths (see Fig. 1). In the construction we treat the nodes of K_δ as one, except for the node that is directly connected to P and similarly the nodes in O .

To enable routing between the n nodes in O , we generalize the construction from Lemma 6. We number the vertices in O with base δ (more precisely, $\max\{2, \lceil \delta \rceil\}$). Then any vertex $v \in O$ corresponds to a unique string $(b_0^{(v)}, \dots, b_{k-1}^{(v)})$ of length $k = \log n / \log \delta$ with each $b_i \in \{0, \dots, \delta - 1\}$. For each $0 \leq i < k$ and $0 \leq j < \delta$ we create a category

$$C_{i,j} := \{v \in O \mid b_i^{(v)} = j\} \cup \{u_j\},$$

where u_1, \dots, u_δ are the vertices of the clique K_δ . Additionally we add a singleton category for each vertex.

Now, every vertex in the clique and O is in at most $k + 1$ such categories. Moreover, we can route between any pair $u, v \in O$, as there is an i with $b_i^{(u)} \neq b_i^{(v)}$, so category $C_{i, b_i^{(v)}}$ allows to route from u to the clique when having v as target.

8 Conclusion and Open Problems

In this paper we presented a improved construction of systems of categories \mathcal{S} that support greedy routing in graphs G . The previous best method uses a membership dimension of $\mathcal{O}((\text{diam}(G) + \log |V|)^2)$, whereas our results show that

$$\mathcal{O}(\text{diam}(G) \log(2|V| / \text{diam}(G)))$$

is sufficient. Besides improved upper bounds we also show much stronger lower bounds than previously known. Our results improve the lower bound from $\text{diam}(G)$ to

$$\Omega(\text{diam}(G) + \log |V| / \log \delta)$$

for graphs of average degree δ . This lower bound is tight for certain graphs. For the restricted class of stars of diameter d and n leaves we showed a lower bound of

$$\Omega((d \log n) / (\log d + \log \log n)).$$

For the case of diameter two the stronger bound

$$\Omega(\log n + \log \log n)$$

holds.

Many open problems remain. Foremost we conjecture that $\mathcal{O}(d \log(2n/d))$

is tight for stars of diameter d , but our lower bound is weaker. Also, the best upper bounds are achieved by taking a spanning tree and constructing categories for it. Is it possible to exploit the properties of the graph better than this?

Empirical studies show that social networks are graphs with a power-law degree distribution and a large clustering coefficient (see e. g. [21] for an overview). The example graphs considered in this paper do not have these properties. It would therefore be interesting to study their consequences in the minimal membership dimension required for routing. Intuitively the membership dimension should also depend on the expansion properties of the graph.

From a sociological point of view it is also interesting to see how natural relaxations of the greedy rule, e. g. allowing nodes to choose a neighbor at random in case of distance ties, influence the bounds.

References

1. Milgram, S.: The small world problem. *Psychology Today* **2**(1) (1967) 60–67 1
2. Travers, J., Milgram, S.: An experimental study of the small world problem. *Sociometry* **32**(4) (1969) pp. 425–443 1
3. Korte, C., Milgram, S.: Acquaintance networks between racial groups: Application of the small world method. *Journal of Personality and Social Psychology* **15**(2) (1970) 101–108 1
4. Bernard, H.R., Killworth, P.D., Evans, M.J., McCarty, C., Shelley, G.A.: Studying social relations cross-culturally. *Ethnology* **27**(2) (1988) pp. 155–179 1
5. Killworth, P., Bernard, H., SOCIOLOGY, W.V.U.M.D.O., ANTHROPOLOGY

- OGY.: The Reverse Small-World Experiment. Defense Technical Information Center (1977) 1
6. Hunter, J.E., Shotland, R.L.: Treating data collected by the "small world" method as a markov process. *Social Forces* **52**(3) (1974) pp. 321–332 1
7. Watts, D.J., Dodds, P.S., Newman, M.E.J.: Identity and search in social networks. *Science* **296**(5571) (2002) 1302–5 1
8. Eppstein, D., Goodrich, M.T., Löffler, M., Strash, D., Trott, L.: Category-based routing in social networks: Membership dimension and the small-world phenomenon (short). *CoRR abs/1108.4675* (2011) 1, 2, 4, 5, 3
9. Finn, G.G.: Routing and Addressing Problems in Large Metropolitan-Scale Internetworks. Technical Report ISI/RR-87-180, Information Sciences Institute (March 1987) 2
10. Kranakis, E., Singh, H., Urrutia, J.: Compass routing on geometric networks. In: IN PROC. 11 TH CANADIAN CONFERENCE ON COMPUTATIONAL GEOMETRY. (1999) 51–54 2
11. Bose, P., Morin, P., Stojmenović, I., Urrutia, J.: Routing with guaranteed delivery in ad hoc wireless networks. *Wireless Networks* **7** (2001) 609–616 10.1023/A:1012319418150. 2
12. Karp, B., Kung, H.T.: Gpsr: greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th annual international conference on Mobile computing and networking. MobiCom '00, New York, NY, USA, ACM (2000) 243–254 2
13. Kuhn, F., Wattenhofer, R., Zhang, Y., Zollinger, A.: Geometric ad-hoc routing: of theory and practice. In: Proceedings of the twenty-second annual symposium on Principles of distributed computing. PODC '03, New York, NY, USA, ACM (2003) 63–72 2
14. Eppstein, D., Goodrich, M.: Succinct greedy graph drawing in the hyperbolic plane. In Tollis, I., Patrignani, M., eds.: Graph Drawing. Volume 5417 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 14–25 10.1007/978-3-642-00219-9_3. 2
15. Goodrich, M., Strash, D.: Succinct greedy geometric routing in the euclidean plane. In Dong, Y., Du, D.Z., Ibarra, O., eds.: Algorithms and Computation. Volume 5878 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2009) 781–791 10.1007/978-3-642-10631-6_79. 2
16. Maymounkov, P.: Greedy embeddings, trees, and euclidean vs. lobachevsky geometry (2006) 2
17. Muhammad, R.B.: A distributed geometric routing algorithm for ad hoc wireless networks. In: Information Technology, 2007. ITNG '07. Fourth International Conference on. (april 2007) 961–963 2
18. Mei, A., Morabito, G., Santi, P., Stefa, J.: Social-aware stateless forwarding in pocket switched networks. In: INFOCOM, 2011 Proceedings IEEE. (april 2011) 251–255 2
19. Kleinberg, R.: Geographic routing using hyperbolic space. In: INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE. (may 2007) 1902–1909 2
20. Chung, F.R.: Separator theorems and their applications. (1989) 5
21. Newman, M.E.J.: The structure and function of complex networks. *SIAM REVIEW* **45** (2003) 167–256 8

Appendix

Here we show the proofs that had to be omitted from the extended abstract due to length limitations. For convenience we repeat the corresponding Lemmas.

Simple Bounds

Lemma. *For any graph G and good system of categories \mathcal{S} we have*

$$\text{memd}(\mathcal{S}) \geq \text{diam}(G).$$

Proof 1. Let $s \in V(G)$, $t \in V(G)$ be a pair of vertices such that $\text{dist}(s, t) = \text{diam}(G)$. Consider the path P that a message from s to t takes according to the greedy routing system. Note that $|P| \geq \text{diam}(G)$. For every edge $(u, v) \in P$ we have $d(v, t) \leq d(u, t) - 1$, and hence

$$d(s, t) \geq |P| \geq \text{diam}(G).$$

By the definition of $d(\cdot)$, this is only possible if $|\text{cat}(t)| \geq \text{diam}(G)$. \square_{\rightarrow}

Lemma. *Take $M(G)$ as the minimal membership dimension needed to route in G . Then*

$$M(G \times H) \leq M(G) + M(H).$$

Proof 2. Let C_G be a minimal category system for G and let C_H be a minimal category system for H . We construct a system of membership dimension $M(G) + M(H)$ for $G \times H$. For every $c \in C_G$, we add $c \times V'$ to the system and for every $c' \in C_H$, we add $V \times c'$ to the system.

We route from (u, x) to (v, y) by first using the first kind of categories to route to (v, x) and then using the second kind of categories to route to (v, y) . \square_{\rightarrow}

Stars

Lemma. *In a star with ℓ leaves, the center node is contained in at least $\log \ell$ categories.*

Proof 3. Let C be any category containing the center and at least one leaf, but not all leaves. Let k be the number of leaves contained in C . If $k \geq \ell/2$ consider the subgraph containing the center and the leaves in C . C cannot be used to route between these leaves. If $k < \ell/2$ consider the subgraph containing the center and the leaves not in C . Since C contains none of the leaves in this subgraph it cannot be used to route between any pair of these leaves.

Let $M(\ell)$ be the minimal number of categories containing the center in a star with ℓ leaves. The paragraph above establishes $M(\ell) \geq 1 + M(\ell/2)$. Thus $M(\ell) \geq \log \ell$. \square_{\rightarrow}

Lemma. *Let k be minimal such that*

$$\binom{k}{\lfloor k/2 \rfloor} \geq \ell.$$

Then $k = \log n + (\frac{1}{2} + o(1)) \log \log n$.

There is a system of categories \mathcal{S} that supports routing in stars with n leaves and diameter 2 having

$$\text{memd}(\mathcal{S}) = k.$$

This bound is tight, i.e., every good system of categories \mathcal{S} for this graph has

$$\text{memd}(\mathcal{S}) \geq k.$$

Proof 4. Let c be the center of the star. For any two distinct leaves i and j there must be categories C and C' such that $C, C' \in \text{cat}(j) \setminus \text{cat}(i)$, $C \in \text{cat}(c)$, and $C' \notin \text{cat}(c)$, as otherwise one cannot route from i to j . Let C_1, \dots, C_k be the

categories containing the center. For every i define a binary vector v_i of length k by

$$(v_i)_k = (i \in C_k) \quad \text{for } 1 \leq k \leq t.$$

The vectors v_i , $1 \leq i \leq \ell$ form an anti-chain in the set of all binary vectors of length k as for every i and j with $i \neq j$, there must be a k with $(v_i)_k = 0$ and $(v_j)_k = 1$. As the maximal size of an anti-chain in the set of all binary vectors of length k is

$$\binom{k}{\lfloor k/2 \rfloor},$$

the lower bound follows.

For the upper bound there are n distinct bitstrings of length k each containing exactly $\lfloor k/2 \rfloor$ ones. Arbitrarily assign the strings to the leaves. We have $k + n$ categories. The latter n categories are singleton sets; they contain one leaf each. The former k categories contain the center and the leaves for which the corresponding bit is one. For any two distinct leaves i and j there is a category C containing j and the center, but not i , because the bitstrings form an anti-chain. We use C to route from i to the center and the singleton category for j to continue to j . Observe that the center is in k categories and each leaf is in $\lfloor k/2 \rfloor + 1 \leq k$ categories.

The equation for k can be derived using Stirling's formula. \square_{\uparrow}

Lemma. *For any star with ℓ leaves, in which every leaf is connected by a path of length d to the center, there is a system of categories \mathcal{S} with membership dimension*

$$\text{memd}(\mathcal{S}) \in \mathcal{O}(d \log \ell).$$

Proof 5. We perform a straightforward extension of the solution for stars of

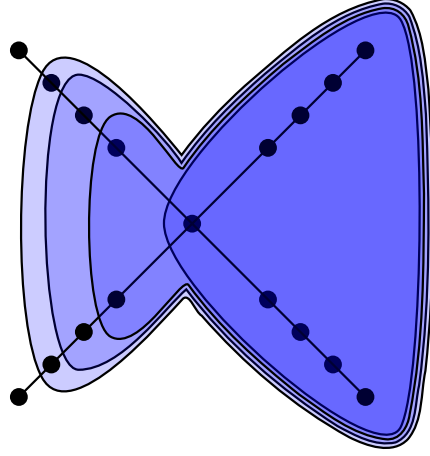


Figure 6. The d copies of a category.

diameter 2. Again every leaf u has a binary number $b(u)$, encoded as categories O_i and Z_i , $1 \leq i \leq \lceil \log \ell \rceil$. Create d copies of these numbers using different categories. Category $O_i^{(k)}$, respectively $Z_i^{(k)}$, contains all nodes on the path from leaves $\{u \mid b_i(u) = 1\}$, respectively $\{u \mid b_i(u) = 0\}$, to the center, as well as all nodes on the remaining paths up to a (graph-)distance k from the center, see Fig. 6.

Additional d categories are needed for every leaf to route from the center down to the leaves, analogous to the singleton categories for leaves in the small star. \square_{\uparrow}

Lemma. *If $4 \ln t \leq \ln \ell$, we have*

$$\sum_k |S_k| \geq \frac{\ell \ln \ell}{32 \ln t}$$

in any covering of Q by t rectangles.

Proof 6. The proof uses a double counting technique. For every i , $1 \leq i \leq \ell$, define a binary vector $v_i \in \{0, 1\}^t$ that

indicates which S_k contain i as

$$(v_i)_k = (i \in S_k), \quad 1 \leq k \leq t.$$

We show that the v_i are pairwise distinct and hence there must be at least ℓ such vectors. Consider any i and j with $i \neq j$. Since (i, j) in Q , there must be a k with $i \in S_k$ and $j \in T_k$. Since $S_k \cap T_k = \emptyset$, we conclude $j \notin S_k$ and hence $(v_i)_k \neq (v_j)_k$.

Clearly the total number of ones in all v_i equals $\sum_k |S_k|$. This number is minimized if there is an h_0 such that all vectors with less than h_0 ones are used, and the remaining vectors contain exactly h_0 ones. Then h_0 must be such that

$$\begin{aligned} \ell &\leq \sum_{h=0}^{h_0} \binom{t}{h} \leq \sum_{h=0}^{h_0} t^h \\ &= \frac{t^{h_0+1} - 1}{t - 1} \\ &\leq t^{h_0+1}. \end{aligned}$$

Thus as $4 \ln t \leq \ln \ell$ by assumption,

$$h_0 \geq \frac{\ln \ell}{\ln t} - 1 \geq \frac{\ln \ell}{2 \ln t}.$$

We split the sum into the vectors that contain less than h_0 ones and the rest. Then the number of ones is at least

$$N := \sum_{h=0}^{h_0-1} \binom{t}{h} h + \left(\ell - \sum_{h=0}^{h_0-1} \binom{t}{h} \right) h_0.$$

We now distinguish cases. If

$$\ell - \sum_{h=0}^{h_0-1} \binom{t}{h} \geq \ell/2$$

then $N \geq (n/2)h_0$. Otherwise there are more than $\ell/2$ vectors with less than h_0 ones and we lower bound N by the

first term. There are two subcases. If $h_0 \leq t/3$ and hence for all $1 \leq h \leq h_0$

$$\binom{t}{h} / \binom{t}{h-1} = \frac{t+1}{h} - 1 \geq 2,$$

we have

$$\begin{aligned} \ell/2 &\leq \sum_{h=0}^{h_0-1} \binom{t}{h} \leq \binom{t}{h_0-1} \sum_{j \geq 0} 2^{-j} \\ &= 2 \binom{t}{h_0-1}, \end{aligned}$$

and hence

$$\begin{aligned} N &\geq \sum_{h=0}^{h_0-1} \binom{t}{h} h \geq \binom{t}{h_0-1} (h_0 - 1) \\ &\geq \frac{\ell}{4} (h_0 - 1) \\ &\geq \frac{\ell \ln \ell}{16 \ln t}. \end{aligned}$$

Finally, if $h_0 > t/3$, we bound

$$\begin{aligned} \sum_{h=0}^{t/4-1} \binom{t}{h} &\leq 2 \binom{t}{\frac{t}{4}-1} \leq \frac{1}{2} \binom{t}{\frac{t}{4}+1} \\ &\leq \frac{1}{2} \sum_{h=t/4}^{h_0-1} \binom{t}{h} \end{aligned}$$

and hence

$$\begin{aligned} \sum_{h=t/4}^{h_0-1} \binom{t}{h} &= \sum_{h=0}^{h_0-1} \binom{t}{h} - \sum_{h=0}^{t/4-1} \binom{t}{h} \\ &\geq \frac{1}{2} \sum_{h=0}^{h_0-1} \binom{t}{h} \\ &\geq \ell/4. \end{aligned}$$

Therefore

$$N \geq \frac{\ell}{4} \cdot \frac{t}{4} \geq \frac{\ell h_0}{16} \geq \frac{\ell \ln \ell}{32 \ln t}.$$

□_→